Probabilistic identication of earthquake clusters using rescaled nearest neighbour distance networks

Geophysical Journal International

Geophys. J. Int. (2019) **217,** 487–503
Advance Access publication 2019 January 18
GJI Seismology

doi: 10.1093/gji/ggz034

Probabilistic identification of earthquake clusters using rescaled nearest neighbour distance networks

Kirsty Bayliss, Mark Naylor and Ian G. Main

School of GeoSciences, University of Edinburgh, James Hutton Road, Edinburgh, EH9 3FE, United Kingdom. E-mail: Kirsty.Bayliss@ed.ac.uk

Accepted 2019 January 17. Received 2019 January 11; in original form 2018 July 23

SUMMARY

Despite the widely recognized importance of the spatio-temporal clustering of earthquakes, there are few robust methods for identifying clusters of causally related earthquakes. Recently, it has been proposed that earthquakes can be linked to their nearest neighbour events using a rescaled distance that depends on space, time and magnitude. These nearest neighbour links may correspond either to causally related event pairs within a clustered sequence or a non-causal relationship between independent events in different sequences. The frequency distribution of these rescaled nearest neighbour distances is consistent with a two-component mixture model where one component models random background events and the other models causally related clusters of events. To distinguish between these populations, a binary threshold has commonly been used to separate the clustered and background events. This has an obvious weakness in that it ignores the overlap of the two distributions and therefore all uncertainty in the event pair classification. It is also restricted so far to treating the two modes as normal distributions. Here we develop a new probabilistic clustering framework using a Markov Chain Monte Carlo mixture modelling approach which allows overlap and enables us to quantify uncertainty in event linkage. We test three hypotheses for the underlying component distributions. The normal and gamma distributions fail to fit the tails of the observed mixture distribution in a well-behaved way. In contrast, the Weibull mixture model is well-behaved in the tail, and provides a better fit to the data. We demonstrate this using catalogues from Southern California, Japan, Italy and New Zealand. We also demonstrate how this new approach can be used to create probabilistic cluster networks allowing investigation of cluster structure and the spatial, temporal and magnitude distributions of different types of clustering and highlight difficulties in applying simple metrics for cluster discrimination.

Key words: persistence, memory, correlations, clustering; Earthquake interaction, forecasting and prediction; statistical seismology.

1 INTRODUCTION

There is a growing need to be able to quantitatively, robustly and reproducibly characterize the properties of earthquake clusters. The motivation has been driven by the growing availability of high resolution earthquake catalogues, the recent focus on operational earthquake forecasting (Jordan *et al.* 2011), including the management of induced seismicity (Bourne *et al.* 2014), and the need to distinguish between different seismic source regions for seismic hazard assessment (Meletti *et al.* 2008; Alvarado *et al.* 2017). Current practice often assumes a mean field statistical description of the clustering properties with a binary separation of dependent and independent events so that catalogues can be 'declustered' to leave only the background component. This practically motivated, binary approach has meant the potentially rich behaviour associated with the overlap of independent and clustered events has remained relatively unexplored. Here we explore how we can relax the binary assumption and hence characterize the differences in the properties of earthquake clusters more accurately.

Broadly, earthquake clustering falls into two subjective categories: mainshock–aftershock and swarm-type sequences. Mainshock–aftershock clusters consist of a large event that triggers further events that decay with time according to the modified Omori law (Utsu *et al.* 1995). Swarm-type sequences are more poorly defined but generally consist of many similar-magnitude events with no obvious 'mainshock'. Recent work (Vidale & Shearer 2006; Enescu *et al.* 2009; Zhang & Shearer 2016) relates the physical properties of the crust in Southern California to the different types of clustering, with swarm-type clustering often associated with fluid movements, high heat flow or aseismic slip.

Recent methods of cluster identification have taken a multidimensional approach to identifying clustering based on empirical laws for earthquake triggering. Baiesi & Paczuski (2004, 2005) developed a space–time–magnitude nearest neighbour distance metric for events, based on (i) the temporal and spatial separation of events, (ii) the Gutenberg–Richter law and (iii) the earthquake hypocentre correlation dimension, suggesting that events with smaller nearest neighbour distances were more likely to be correlated. This involves the product of a rescaled separation distance respecting the fractal geometry of earthquake hypocentres, the interevent time, and the fact that large events are more likely to trigger more and later events. Zaliapin & Ben-Zion (2013a,b, 2016a) developed this idea further by attempting to separate the bimodal nearest-neighbour distribution observed by Zaliapin *et al.* (2008) into two constituent components: background events with larger rescaled nearest neighbour distance, and clustered events with smaller rescaled nearest neighbour distance. They applied a binary threshold determined by the crossover point of two normal distributions, fitted with an expectation maximization approach to the distribution of rescaled nearest neighbour distances and constructed deterministic networks or chains of linked events to identify clusters, and highlight differences in clustering styles. The binary approach means the links are identified as independent or dependent with a probability one. They then proposed a topological approach to separating the identified clusters into different types, specifically into mainshock–aftershock and swarm type clusters and related these to underlying physical properties of the crust.

This nearest neighbour method has subsequently been applied or modified to investigate induced seismicity (Schoenball *et al.* 2015; Zaliapin & Ben-Zion 2016b; Schoenball & Ellsworth 2017), the background rate of events (Gentili *et al.* 2017), swarm sequences (Ruhl *et al.* 2016; Zhang & Shearer 2016), Italian earthquake clusters (Peresan & Gentili 2018) and as a comparison cluster identification technique (Moradpour *et al.* 2014; Reverso *et al.* 2015; Maghsoudi *et al.* 2016). However, the binary threshold approach ignores the overlap in the two distributions inherent in a mixture model and therefore ignores uncertainty in the links. This is particularly problematic for events with nearest neighbour distance close to the threshold value, which itself will inevitably be in the area of highest uncertainty of event classification given the nature of the model. The threshold itself is also highly dependent on the assumptions made in the mixture model fit, for example the choice of two normal distributions. Any associated bias in the fit would then propagate through to the threshold selection and thus the constructed clusters. In addition location errors could lead to a link crossing the threshold and changing its type from causal to non-causal, or vice versa (Zaliapin & Ben-Zion 2015).

Here we develop a new probabilistic approach to categorizing nearest neighbour event pairs in order to construct distinct earthquake clusters in the form of a network or chain of links with a probability of being causal between zero and one. We implement a Markov Chain Monte Carlo (MCMC) mixture model as an alternative to the expectation maximization approach employed by Hicks (2011) and Zaliapin & Ben-Zion (2013a,b, 2016a). This allows us to sample not one but many mixture fits to the distribution to highlight the uncertainty in the model fit, and how this propagates into the categorization process. We test three hypotheses for the underlying component distributions—normal distributions assumed by Zaliapin & Ben-Zion (2013a), gamma distributions and Weibull distributions as alternates. We apply the three models to earthquake catalogue data from Southern California. The Weibull model is clearly preferred as a better fit to the data, with well-behaved tails, and we demonstrate that it is also a good fit to independent data sets from Japan, Italy and New Zealand. We then identify clusters of events where each link has an associated probability of causality, and examine the nature of the clusters. Finally, we demonstrate how this link uncertainty can then be propagated to the resulting cluster statistics.

2 THEORY

2.1 Nearest neighbour distances

Zaliapin & Ben-Zion (2013a) proposed a space–time–magnitude nearest neighbour distance metric based on earlier work by Baiesi & Paczuski (2004), where for each event j the rescaled distance to each previous event i

$$\eta_{ij} = t_{ij}(r_{ij})^{d_f} 10^{-bm_i}, \tag{1}$$

where t_{ij} is the interoccurrence time in years between events i and j and r_{ij} is the epicentral distance in kilometres. m_i is the magnitude of the potential parent event, b is the Gutenberg–Richter b-value and d_f is the earthquake epicentre correlation dimension (Kagan 2007) describing the spatial distribution of earthquakes. The nearest neighbour parent is then the event which has the minimum distance from an event j, such that

$$\eta_j = \min(\eta_{ij}). \tag{2}$$

The metric η_j can be considered a rescaled nearest-neighbour distance, and will be referred to as such in this paper to distinguish from a spatial or hyperspatial distance in the text to follow. For most of this paper we use the relocated Southern California earthquake catalogue from Hauksson *et al.* (2012) (available from http://scedc.caltech.edu/research-tools/downloads.html) for 1981–2011. This includes 111 981 events with $M \geq 2.0$. Following current practice, we assume a fixed value of $b = 1$ in this work (e.g. Zaliapin & Ben-Zion (2013a)), close to calculated values reported in the literature for Southern California (e.g. Kamer & Hiemer 2015). This choice leads to some circularity in the method since fixing b implies that we do not expect it to vary between clusters of different style. This may not always be the case, for example in volcano-tectonic settings. A d_f value of 1.6 is used throughout this work as suggested by Zaliapin & Ben-Zion (2013a, 2016a). Zhang & Shearer (2016) tested several variations of this value in Southern California and found that the cluster identification was not greatly affected by alternate choices, though Vasylkivska & Huerta (2017) and Peresan & Gentili (2018) both found that in some areas this could have a significant impact on the binary separation threshold for causally related events.

The inclusion of magnitude in such an analysis allows larger events to be linked with events that occur at greater temporal or spatial distance than smaller events. This is a big advantage over purely spatiotemporal clustering methods or analysis, because it acknowledges the greater effect of larger events.

The frequency distribution for the rescaled nearest neighbour distance η_j in Southern California is shown in Fig. 1(a). It confirms that the distribution is bimodal. We can also decompose the distribution into two dimensions by splitting our η_j into a rescaled time and a rescaled distance component such that

$$t_j = t_{ij} 10^{-0.5bm_i}, \tag{3}$$

$$r_j = (r_{ij})^{d_f} 10^{-0.5bm_i}, \tag{4}$$

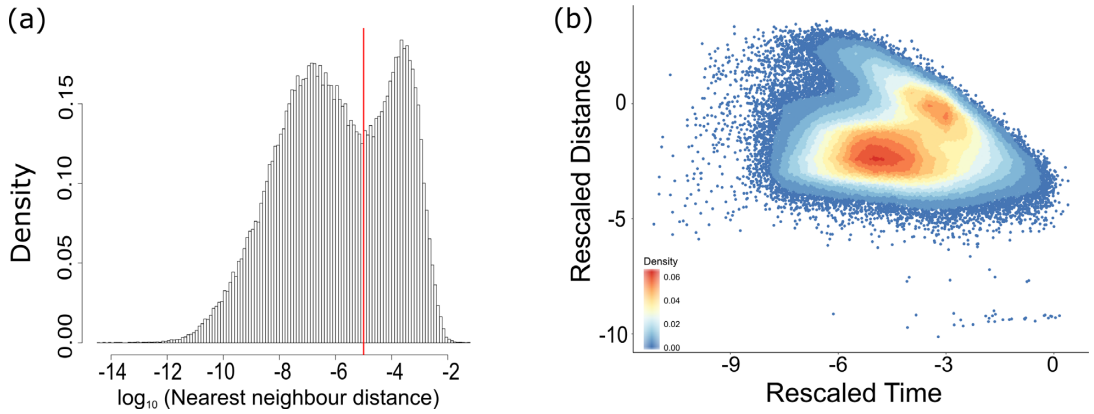

Figure 1. (a) Histogram of rescaled nearest neighbour distances η_j for the Southern California M2+ catalogue 1981–2011 with the binary threshold suggested by Zaliapin & Ben-Zion (2013a) marked by the vertical red line. (b) Rescaled time (t_j) and rescaled distance (r_j) components (see eqs 3–5) for the same catalogue confirming the bimodal distribution of nearest neighbour distances in two dimensions.

$$\eta_j = t_j r_j, \tag{5}$$

where we have already calculated the most likely parent event i based on the rescaled nearest-neighbour distance η_j. The frequency distribution for these two components is shown in Fig. 1(b). These are also bimodal. One mode is at larger rescaled times and distances, which Zaliapin & Ben-Zion (2013a) demonstrate is equivalent to a Poisson process of independent background events. The other is at much smaller times and distances, indicating event pairs that are more clustered than expected for a Poisson process. Hicks (2011) confirmed this bimodality also exists in catalogues from New Zealand and Japan, so we have also applied our method to these areas.

Baiesi & Paczuski (2004, 2005) suggest that events with the smallest nearest neighbour distance are less likely to have occurred randomly, and thus are the most highly correlated events. Zaliapin & Ben-Zion (2013a) extended this by applying a normal mixture model to the bimodal nearest neighbour distribution and applied a threshold determined by the crossover point of the two distributions to split the two components into background and clustered events. Moradpour *et al.* (2014) proposes that separating the 2-D rescaled time-distance distributions (Fig. 1b) is a better approach than separating the 1-D distribution (Fig. 1a). This is a popular alternative to the 1-D separation (e.g. Davidsen *et al.* 2015; Maghsoudi *et al.* 2016, 2018).

If we assume that the two components of the distribution arise by different mechanisms then the overlap of the two components must also be important in describing the combined distribution. The fixed binary threshold of Zaliapin & Ben-Zion (2013a) neglects this. Therefore we propose an MCMC approach (e.g. Gallagher *et al.* 2009) to the mixture modelling as outlined below which will allow us to investigate this overlap and take a probabilistic approach to categorizing events as dependent or independent, and hence part of a causally related cluster or not.

2.2 Mixture modelling

Mixture models are used to model distributions which are composed of, or can be approximated by, a linear sum of component distributions. Here we explore mixtures of different distribution types to identify a robust approach to modelling the distribution of nearest neighbour distances. We require two components in the mixture, one for the independent or background events and another for the causally related or triggered events that form clusters.

Normal distributions are most commonly used for the components of mixture distributions and there are many standard tools available to perform this analysis. Mixtures of other types of distribution are much less commonly employed. We consider three different distribution types to fit the nearest neighbour distance distribution: the standard normal mixture applied to the log of the rescaled nearest neighbour distance distribution making it log-normal in linear space, a Gamma mixture and a Weibull mixture. There are theoretical reasons for the choice of the Gamma and Weibull distributions to the background component of the rescaled nearest neighbour distance as discussed in the following sections, and the (log) normal distribution is used for comparison due to its use in the existing literature. We restrict the analysis in this paper to pairs of the same component mixture models, given the lack of theoretical models describing the rescaled nearest-neighbour distance behaviour of clustered events.

A mixture model consists of k component distributions, where the component distributions $F(x)$ are each described by a vector of parameters $\mathbf{a}_j$ and the contribution of each component distribution is described by some weighting parameter w, where the sum of the elements of w over all components will be equal to unity, such that

$$P(x \mid \mathbf{a}, \boldsymbol{w}, k) = \sum_{j=1}^{k} w_j F(x \mid \mathbf{a}_j), \tag{6}$$

where $\mathbf{a}$ is a matrix containing the parameters for each component $\mathbf{a}_j$. Thus, the conditional probability of some value x, given the parameters of the component distributions and their weighting, is equal to the sum of the weighted probabilities of the component distributions.

Here we restrict the analysis to two components ($k = 2$) defining a background and clustered distribution of rescaled nearest neighbour distances η_j. In order to model these components we must choose a suitable distribution for each component of the mixture. We can then use MCMC sampling to identify appropriate values for the distribution parameters $\mathbf{a}_j$ and for the weighting parameter w_j, as well as for the uncertainties in the underlying parameters and the resulting frequency distribution.

The MCMC method relies upon a Bayesian hierarchical model to identify a number of possible model solutions by sampling from the posterior distributions, which are generated using Bayes theorem.

An MCMC approach to mixture modelling thus aims to find suitable posterior values for all the model parameters. We take a Gibbs sampling approach to mixture modelling as described by Diebolt & Robert (1994) which samples each parameter posterior conditional on the values of the other parameters. We assign to each parameter of interest a prior distribution which is described by some hyperparameters. This method also includes a step which assigns each data point a variable Z_i which describes if the event belongs to one distribution or the other, which is calculated based on the other parameter values at the current iteration such that

$$P(Z_i = j \mid k, \boldsymbol{w}) = w_j, \tag{7}$$

$$(x_i \mid k, Z_i = j) \sim F(x_i \mid \mathbf{a}_j), \tag{8}$$

where eq. (8) describes how the subset of x assigned variable $Z_i = j$ is described by parameters $\mathbf{a}_j$. Our data x in the following distributions are the rescaled nearest neighbour distances η_j.

Gibbs sampling samples from the conditional posteriors of the parameters making use of conjugate pairs—if the prior and the posterior distribution are conjugate and the posterior is of the form of a known distribution, the posterior can be sampled directly as the hyperparameters of the prior are related to those of the posterior. It does this for each parameter in the model, conditional on the other parameter values. As such, if conjugate priors are available, an acceptance or rejection step is not required and such a model is therefore more straightforward to implement than other sampling methods. For mixtures of data where the posterior is not a known distribution, an acceptance ratio must be determined so that new proposed values can be accepted or rejected. For this we use a Metropolis step which makes a small perturbation to the existing parameter value drawn from a normal distribution.

2.2.1 Normal mixtures

Normal mixture models have been applied by several authors (e.g. Zaliapin & Ben-Zion 2013a, 2016a) to analyse the interevent distance distributions. This is primarily due to the ease of implementation of a normal mixture and a reasonable first order fit to the data. There is no underlying theoretical reason to prefer normal distributions. The normal mixture model was implemented in R (R Core Team 2018) using OpenBUGS (Lunn *et al.* 2009) and the R2OpenBUGS package (Sturtz *et al.* 2005).

For a mixture of two normal distributions the probability density function is

$$p(x \mid \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{w}, k) = \sum_{j=1}^{2} w_j \; N(x \mid \mu_j, \sigma_j^2), \tag{9}$$

where μ is the mean of the distribution and σ^2 is the variance. It is often more useful in this context to consider the precision $\tau = \frac{1}{\sigma^2}$ which allows the use of conjugate priors for μ and τ. For a normal distribution, choosing a normal prior for the mean will ensure a normal distribution for the posterior of the mean. Once a prior distribution has been chosen, hyperparameters for this distribution are also specified. In the case of a Gaussian conjugate prior, the hyperparameters would be μ_0 and τ_0. A conjugate prior also exists for the precision parameter of a Gaussian distribution: a gamma distribution with hyperparameters α and β, which gives a gamma posterior distribution.

The mixture weighting $\boldsymbol{w}$ is determined through a Dirichlet distribution. This gives our prior distributions as

$$\mu_j \mid k \sim Normal(\mu_0, \tau_0), \tag{10}$$

$$\tau_j \mid k \sim Gamma(\alpha, \beta), \tag{11}$$

$$\boldsymbol{w} \mid k \sim Dirichlet(\phi_1, \phi_2), \tag{12}$$

where $j = 1, 2$ and we can set $\phi_1 = \phi_2 = 1$ to give a uniform prior on the weights. The posterior densities can be approximated by combining the likelihood of the mixture distribution with the selected prior distributions. For a normal distribution this gives conditional posterior densities of

$$\mu_j \mid x, k, \boldsymbol{Z}, \boldsymbol{\tau} \sim Normal\Big(\frac{\mu_0\tau_0 + \tau_j \sum_{i=1}^{n} x_i}{n\tau_j + \tau_0}, n\tau_j + \tau_0\Big), \tag{13}$$

$$\tau_j \mid x, k, \boldsymbol{Z}, \boldsymbol{\mu} \sim Gamma\Big(\alpha + \frac{n}{2}, \beta + \frac{\sum_{i=1}^{n}(x_i - \mu_j)^2}{2}\Big), \tag{14}$$

$$\boldsymbol{w} \mid x, k, \boldsymbol{Z}, \boldsymbol{\mu}, \boldsymbol{\tau} \sim Dirichlet(\phi_1 + n_1, \phi_2 + n_2), \tag{15}$$

where n_j in the Dirichlet() is the number of data assigned to component j.

For a two component normal mixture, there are five free parameters: two means, two precisions and a weight. The range of support of the normal distribution allows for the normal mixture to be fitted to the log of the nearest-neighbour distribution, but the Gamma and Weibull distributions are strictly limited to positive values and so are fitted to the original η_j values and then transformed to log space in Fig. 2 for ease of visualization.

2.2.2 Gamma mixtures

The next most obvious choice is a mixture of gamma distributions. The motivation of the gamma mixture is derived from Touati *et al.* (2009) who demonstrated that the earthquake interevent time distribution can be modelled using a gamma-exponential mixture. If we consider the background events to be random in time with a constant rate, an exponential distribution of interevent times should be anticipated. The gamma distribution models the triggered events.

This Gamma-Exponential mixture is just a special case of a two component gamma mixture where the shape parameter of one of the distributions $\alpha = 1$, so a mixture of two gamma distributions was implemented using R code modified from Mohammadi *et al.* (2013) for a two component gamma mixture model, which makes use of functions in the gtools package (Warnes *et al.* 2015). Restricting the shape parameter of the background distribution in this way, however, resulted in a very poor fit to the nearest-neighbour distribution, and this assumption was relaxed in further modelling and in the results demonstrated below.

The probability density function for a mixture of two gamma distributions, each with shape parameter α and rate parameter β is then

$$p(x \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{w}, k) = \sum_{j=1}^{2} w_j \; Gamma(x \mid \alpha_j, \beta_j). \tag{16}$$

The rate parameter β is sampled from a gamma conjugate prior $Ga(\alpha_0, \beta_0)$. The α parameter is also assigned a gamma prior $\alpha \sim$

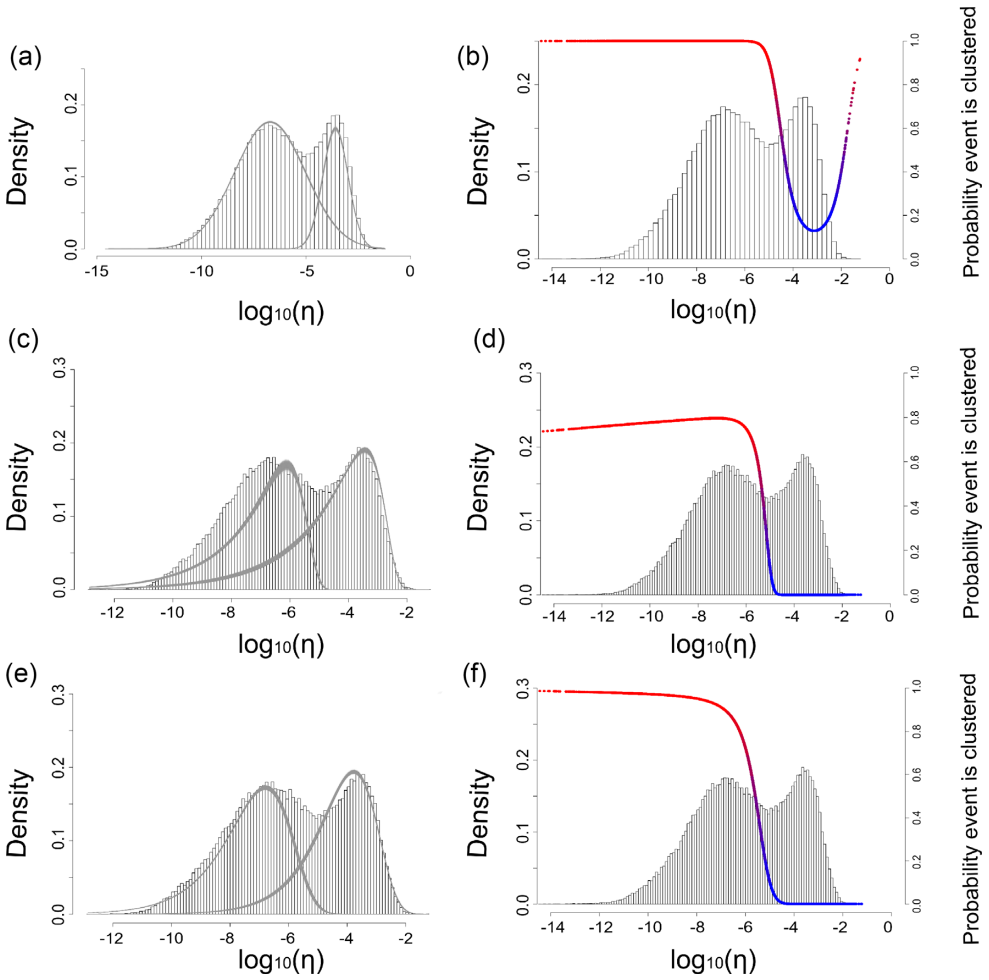

Figure 2. Two component mixture modelling of the Southern California earthquake catalogue for $M > 2$ using MCMC. The top row shows the results for the normal mixture model, the middle row shows the gamma mixture and the bottom row the Weibull mixture. The left hand figures use the posteriors to plot 100 samples of each mixture distribution in grey on top of the rescaled nearest neighbour distribution (shown as a histogram) to illustrate the uncertainty in the fit. The right figures show the same histograms with the conditional probability that an event with given η_j belongs to the clustered distribution given the mean parameter posterior values for both component distributions. The colour of the line is also used to highlight this with red events having cluster probability = 1 and blue events with cluster probability = 0.

$Ga(\mu, \nu)$ but in this case the posterior is not of known form and so a Metropolis step is required to sample from the posterior by making slight perturbations to the previous value drawn from $N(0, \zeta)$

The conditional posterior distributions are then:

$$\boldsymbol{w} \mid x, k, \boldsymbol{Z}, \boldsymbol{\alpha}, \boldsymbol{\beta} \sim Dirichlet(\phi_1 + n_1, \phi_2 + n_2), \tag{17}$$

$$\beta_j \mid x, k, \boldsymbol{Z}, \boldsymbol{\alpha} \sim Gamma(\alpha_0 + n_j\alpha_j, \beta_0 + \sum x_i), \tag{18}$$

$$f(\alpha_j \mid x, k, \boldsymbol{Z}, \boldsymbol{\beta}) \propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\right)^{n_j} \left(\prod x_i\right)^{\alpha_j} \alpha_j^{\mu-1} e^{-\nu\alpha_j}, \tag{19}$$

where $i = 1...n$, where n is the number of data points. With the acceptance ratio for the new α values calculated using the equation:

$$Pr(\alpha, \alpha^*) = \min(1, r), \tag{20}$$

where r is defined as

$$r = \left(\frac{\Gamma(\alpha)}{\Gamma(\alpha^*)}\right)^n \left(\prod \beta x\right)^{\alpha^*-\alpha} \left(\frac{\alpha^*}{\alpha}\right)^{\mu-1} e^{\nu(\alpha^*-\alpha)} \frac{q(\alpha^*)}{q(\alpha)}. \tag{21}$$

The acceptance ratio is calculated by comparing posterior densities of the current and updated α distribution and multiplying this by the ratio of updated to current values of the proposal function. In this case a Gaussian proposal distribution has been used. The symmetry of the Gaussian distribution would mean this ratio was 1 and the ratio is simply that of posterior densities of α. However as the proposals are for a distribution with uneven support (α must be greater than 0), it is necessary to account for this in the acceptance ratio. The parameter $q()$ is the ratio of normalizing constants that are not cancelled here due to their dependence on α and $\alpha*$—as we are using a Gaussian distribution this is simply the Gaussian CDF at α/ζ and $\alpha*/\zeta$. This step has been included for mathematical completeness, in practice values less than 0 were never encountered.

For a two component gamma mixture, there are five free parameters: two shape parameters, two rate parameters and a weight. Forcing a gamma-exponential mixture reduces this to four free parameters since the shape parameter for the exponential component is identically one.

2.2.3 Weibull mixtures

The Weibull distribution was suggested by Zaliapin *et al.* (2008) as the distribution of nearest neighbours resulting from an n-dimensional time-stationary space-homogeneous Poisson process. Hicks (2011) demonstrates that for an n-dimensional sphere the distribution of nearest neighbours from an n-dimensional Poisson process would follow a Weibull distribution, and Zaliapin & Ben-Zion (2013a) demonstrate that the Weibull CDF is a good approximation to the cumulative distribution of nearest neighbour distances. However, they did not formally implement it within a mixture model. The Weibull distribution is also closely related to the exponential distribution, with the exponential being a special case of the Weibull distribution with scale parameter $a = 1$.

We use the following Weibull probability density:

$$f(x) = a\theta x^{a-1} \exp(-\theta x^a), \tag{22}$$

for a shape parameter a and a scale parameter θ, as in Marín *et al.* (2005). This gives the probability density function for a mixture of two Weibull distributions as

$$p(x \mid \boldsymbol{a}, \boldsymbol{\theta}, \boldsymbol{w}, k) = \sum_{j=1}^{2} w_j \ \text{Weibull}(x \mid a_j, \theta_j). \tag{23}$$

Choosing this parametrization of the Weibull distribution allows for the use of a gamma conjugate prior for the scale parameter $\theta_j \sim Gamma(\alpha_\theta, \beta_\theta)$, and we use a gamma prior for a such that $a \sim Ga(\alpha_a, \beta_a)$ which again requires a Metropolis step. The conditional posteriors are then:

$$\boldsymbol{w} \mid x, k, \boldsymbol{Z}, \boldsymbol{\theta}, \boldsymbol{a} \sim Dirichlet(\phi_1 + n_1, \phi_2 + n_2), \tag{24}$$

$$\theta_j | x, k, \boldsymbol{Z}, \boldsymbol{a} \sim Gamma\left(n_j + \alpha_\theta, \beta_\theta + \sum x_i^{a_j}\right), \tag{25}$$

$$f(a_j | k, x, \boldsymbol{Z}, \boldsymbol{\theta}) \propto a_j^{n_j+\alpha_a-1} \exp \times \left\{-a_j\left(\beta_a - \sum \log x_i\right) - \theta_j \sum x_i^{a_j}\right\}, \tag{26}$$

The acceptance ratio for new a values is then

$$Pr(a, a^*) = \min(1, r), \tag{27}$$

where r is now

$$r = \left(\frac{a^*}{a}\right)^{n+\alpha_a-1} \exp \times \left(\left(\beta_a - \sum(\log(x_i))\right)(a - a^*) + \theta \sum (x_i^a - x_i^{a^*})\right) \frac{q(a^*)}{q(a)}, \tag{28}$$

and again we include a step to account for the uneven support of the distribution in the Metropolis step.

For a two component Weibull mixture, there are five free parameters: two shape parameters, two scale parameters and a weight.

3 RESULTS

3.1 Three mixture models applied to the Southern California earthquake catalogue

We have applied the Normal, Gamma and Weibull mixture models to the nearest neighbour distance distribution derived from the relocated Southern California earthquake catalogue from Hauksson *et al.* (2012) (available from http://scedc.caltech.edu/research-tools/downloads.html) for 1981–2011. This includes 111 981 events with $M \geq 2.0$ (Fig. 2).

When judging the quality of the fitting, we are interested in two factors. The first is how well the mixture model describes the distribution of the rescaled nearest neighbour distance η_j. The second is whether the tails of the probability distribution of an event pair being background or clustered are well behaved, i.e. that the probability of a pair of events being clustered tends to 0 at large η_j and tends to 1 at short η_j. This second criterion is illustrated in the right hand column of Fig. 2, which shows the probability of an event belonging to the clustered distribution. The colour of the curve indicates the probability of an event being clustered, such that blue events have 0 probability of being clustered and red events have a probability of being clustered of one.

Visually, the Normal mixture is a reasonable fit to the underlying rescaled nearest neighbour distribution to first order and results in small uncertainties in the fit as shown by the narrow range of possible fits shown in grey (Fig. 2a). Hence it appears to describe the clustered component very well. However there is a bias. The normal distribution overestimates the number of events with the largest η_j, expecting more events at greater η_j than are present in the catalogue. Our analysis showed that this problem is particularly troublesome for smaller data sets with a higher minimum magnitude cut-off. The slow decay of the tails at increasing distance from the mean of each component is problematic for both background and clustered components of the distribution, with the tails of the clustered component decaying at a much slower rate than the tails of the background component. This causes significant overlap of the distributions for events with large η_j and suggests significant uncertainty in links at these distances. This overlap of the tails at the largest distances causes significant problems. The conditional probability of an event being clustered increases as a consequence of the overlapping tail behaviour (Fig. 2b). Due to the shape and slow-decaying tail of the clustered distribution compared to background, the probability of events belonging to the clustered distribution does not tend to 0 with increasing η_j and in fact increases at the largest η_j. It is unphysical for the largest rescaled nearest neighbour distances to be associated with the triggered component and unlikely that events at significant η_j would be the most uncertain. Hence we can reject the normal mixture model for the distribution of rescaled nearest neighbour distances η_j.

The Gamma mixture model Fig. 2(c) is a better fit for the largest η_j events, but this time the fit to the clustered component at low η_j is poor. Moreover the tail of the Gamma fit to both component distributions is long and slow to decay, which results in more significant overlap of the distributions and a greater area of uncertainty where it is unclear which distribution individual event pairs should be associated with. This is reflected in the probability graph, as the probability that an event belongs to the cluster distribution levels off at around 0.8, and then slightly decreases for the smallest η_j, again due to the overlapping of the distribution tails. This behaviour at short interevent distances is also unphysical—we would expect events with the smallest rescaled nearest neighbour distances to be almost certain to be causally related, with probability close to or identical to 1.

The Weibull mixture provides the best fit to the two component distributions and the distribution as a whole. There is less spread in the resulting distributions for the Weibull mixture (Fig. 2e) than that of the Gamma mixture (Fig. 2c), indicating less uncertainty in the fit. The tails of the Weibull distribution are faster to decay than the gamma, reducing the area of overlap of the two components for

Table 1. Log-likelihoods for Weibull and Gaussian mixture fits to Southern California $M2+$, Japan $M4.5+$ 1976–2016, Italy $M2+$ 2006–2016 and New Zealand $M2.5+$ 2007–2017 rescaled nearest neighbour distances.

Region	Weibull	Gaussian	Gamma
Southern California	1 204 592	1 205 102	1 200 675
Japan	173 826	173 992	–
Italy	245 878	245 683	–
New Zealand	801 123	798 274	–

this catalogue. This area of overlap has now been shifted to more intermediate distances compared to the normal mixture. The cluster probabilities are well behaved at both large interevent distances where it decays to 0 and at short interevent distances where it tends to 1.

The log-likelihoods for each of the mixture models were calculated and are shown in Table 1. The gamma mixture performs the worst in terms of log-likelihood, but the Weibull and normal distributions perform similarly well. The log-likelihoods relate to the total mixture fit, and not the performance of the individual components. Given the similarity in the resulting likelihoods, we impose the requirement that the clustering probabilities resulting from the fits is consistent with expected physical behaviour.

For each of these fits we present the parameter posterior and prior distributions in Fig. 3. The Gamma mixture priors are clearly not ideal, and broader priors may have led to faster convergence. The other prior distributions are intentionally broad so as not to unnecessarily restrict possible solutions, and the resulting posteriors show a very limited range in comparison. This low uncertainty in the fits is achieved by using larger data sets—smaller data sets will have a larger uncertainty and broader posterior distributions. This is demonstrated in Fig. 4 which splits the California catalogue into one year segments and calculates the Weibull mixture fit for each segment. The posteriors of the individual years are much more variable than for the full catalogue due to the smaller data set. Many of the fits in this figure are also clearly biased from the larger data set results by the distribution of individual years—this is highlighted particularly in years which have experienced large earthquakes which leads to an increase in clustered events and a markedly different total nearest neighbour distribution compared to years without large event sequences. This is also the cause of the large variation in the weighting parameter w. This suggests that there may be considerable differences in linked events if different subsections of the same catalogue are used. The mean posterior values for each parameter are used to calculate the probability fits in Fig. 2, though in practice we use the full parameter posterior distributions in the construction of the probabilistic networks as outlined below.

All of these mixture models highlight the potentially significant overlap between the clustered and background components. This demonstrates a degree of realistic uncertainty in how event pairs should be classified. It is very clear from these plots that the practice of using the minimum between the two distributions to separate the background and clustered components (as in Zaliapin *et al.* (2008)) is a significant oversimplification when characterizing the background and clustered components. Our method accounts for the allocation uncertainty arising from the overlap (illustrated in the right-hand column of Fig. 2). This allows us to develop and

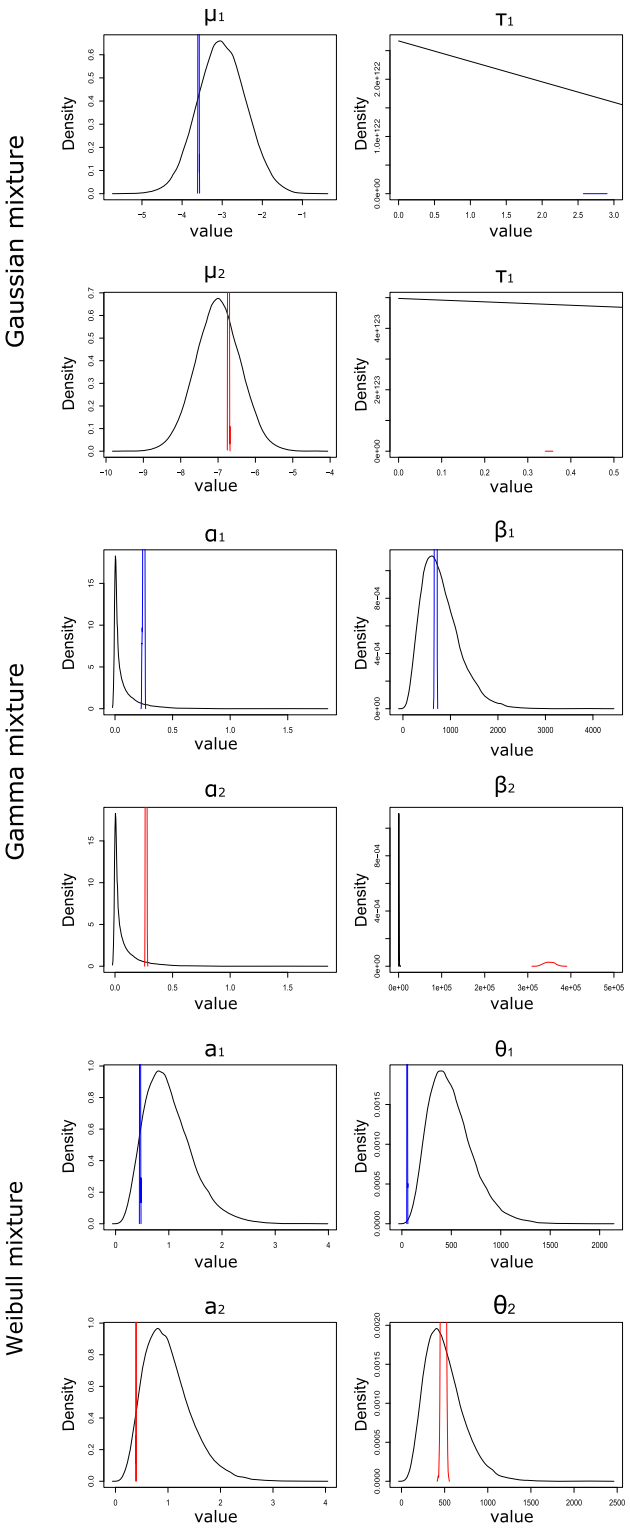

Figure 3. Posterior distributions for the mixture model parameters (coloured blue for background components and red for the clustered component distribution) for the normal (top, see eq. 9), gamma(middle, see eq. 16) and Weibull (see eq. 23) mixtures, with the prior distributions shown in black.

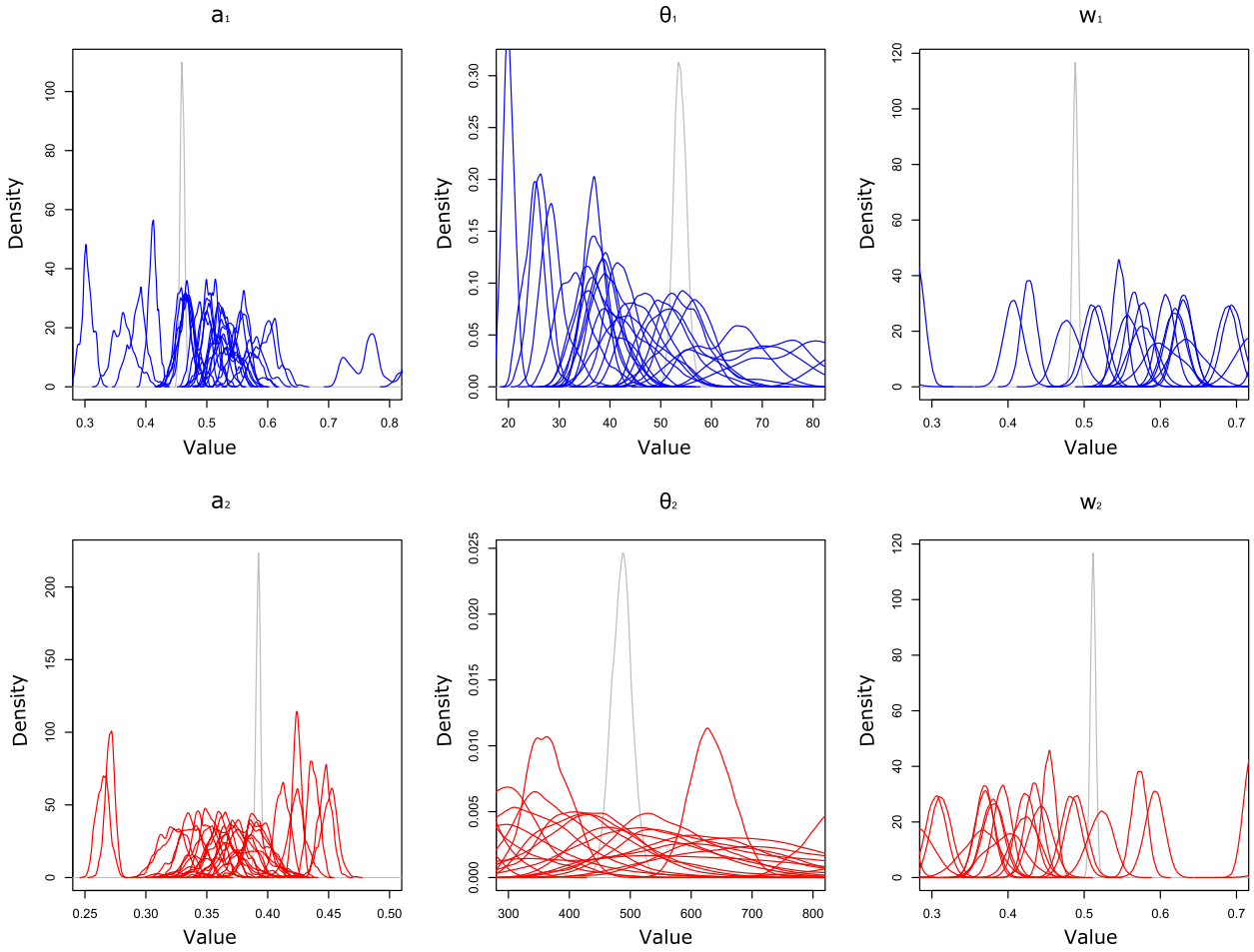

Figure 4. Posterior distributions for the parameters of the Weibull fit for yearly Southern California catalogues, with the posterior for the total catalogue shown in grey. Blue represents the parameters of the background component and red the parameters of the clustered component.

apply a new method for probabilistically categorizing event pairs as causally related or not, as presented in Section 3.3. First, we demonstrate the suitability of the Weibull mixture on three different catalogues.

3.2 Application of Weibull mixture modelling to Japanese, New Zealand and Italian catalogues

We now apply the Weibull mixture model to three further catalogues to demonstrate the opportunities and some limitations of the method. Fig. 5 shows samples taken from the posteriors for the Japanese, Italian and New Zealand catalogues.

The Japan catalogue consists of 16 744 events of $M \geq 4.5$ between 1976 and 2016 as taken from the USGS catalogue (earthquake.usgs.gov/earthquakes for the area covered by latitude 20–50, longitude 129–150 between 4/10/1976 and 4/10/2016). The New Zealand data includes 84984 events of magnitude $M \geq 2.5$ between 1/1/2007 and 1/1/2017, available from GeoNet (https://www.geonet.org.nz/data/types/eq_catalogue). The Italian data set is from INGV (at http://cnt.rm.ingv.it/,) for $M \geq 2.0$ events between 12/10/2006 and 12/10/2016.

The Weibull mixture performs well on all three catalogues. In each case we see that the background peak is more prominent than the clustered one. This is less obvious in the Southern California catalogue (Fig. 1a) or shorter data sets which may have a larger component of short-term clustering, particularly if windowed around larger events. This suggests that the choice of catalogue or time window for analysis will have a significant impact on the resulting mixture model fits, and therefore on the links themselves. The Italian catalogue has a much larger cluster component than the New Zealand or Japanese catalogues, both of which have a significant proportion of total events in the background distribution. This is further confirmed by studying the resulting weights of each component distribution. While the Southern California and Italian catalogues have a roughly 50–50 split of clustered and background events, the Japanese fits suggest 69 per cent of events are in the background distribution and the New Zealand fits suggest 83 per cent of events are in the background distribution. In the case of the Japanese catalogue, this may be due in part to the higher magnitude cut-off, which would exclude many more smaller magnitude aftershocks. The Japanese earthquake catalogue has the greatest uncertainty in the Weibull fits, as shown by the wider confidence band from superposing the MCMC fits in Fig. 5(a) compared to Figs 5(b) and (c). This is a result of the smaller catalogue which leads to greater uncertainty in the fit.

A log-likelihood comparison for normal and Weibull mixtures for each of these data sets is shown in Table 1. The table shows that in the New Zealand and Italian examples, the Weibull model performs better than the normal, but that the opposite is true in Southern

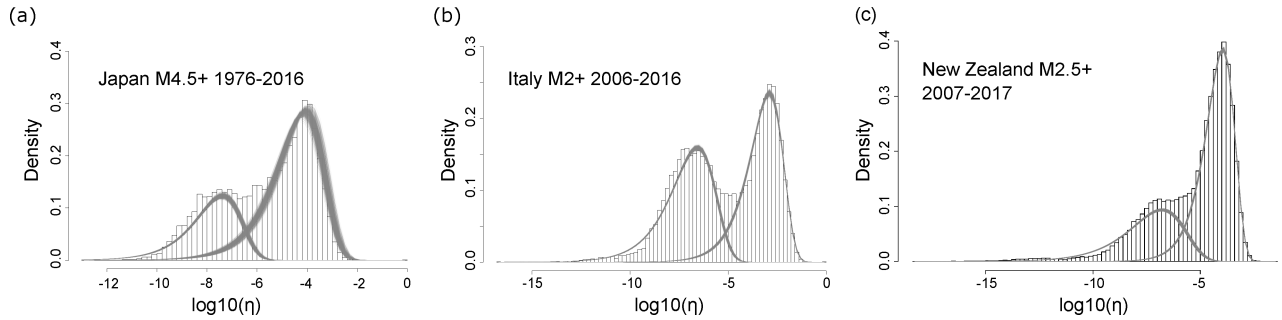

Figure 5. Nearest neighbour distance distribution histograms for (a) Japan 1976–2016 $M \geq 4.5$, (b) Italy 2006–2016 $M \geq 2.0$ and (c) New Zealand 2007–2017 $M \geq 2.5$.

California and Japan. In all cases, the log-likelihoods are similar for either model. The shape of the clustered distribution varies in different regions and over different time periods, as demonstrated by the posteriors of the yearly fits in Fig. 4. It is therefore important to construct a model that fits the background component, which is demonstrably more similar in different areas, than to attempt to perfectly fit the clustered distribution, which varies between different data sets. Furthermore, as demonstrated in the right hand column of Fig. 2, the Weibull distribution is the mixture which results in clustering probabilities that are consistent with current understanding of earthquake clustering behaviour, with events which are closer together in some space-time sense more likely to be clustered than events with larger rescaled nearest neighbour distances.

In all three cases the clustered distribution is stretched to the smallest η_j due to events at very small rescaled nearest neighbour distances. In the Japanese earthquake catalogue there are only a few of these events, and these are events that occur very close in time or space or have a very large magnitude parent. In both the New Zealand and Italian catalogues these small events cause significant tails towards the smallest η_j values. In Fig. 6 we examine how this arises by separately considering the spatial (b, h), temporal (c, i) and magnitude scaling (d, j) components of the smallest nearest neighbour distance pairs, those with $\eta_j < 10^{-9}$ in the Italian case (982 of 27 939 events) and $\eta_j < 10^{-11}$ in the New Zealand data (1398 of 84 984 events) as shown by the dashed lines in Figs 6(a) and (g). These thresholds were chosen by examining the rescaled nearest neighbour distributions to see where the tails might be expected to decay, as occurs in other mixture fits without the very small rescaled nearest neighbour distance events. We also show how these event pairs with small η_j occur within the larger catalogue. The events at very small rescaled nearest neighbour distances defined by these thresholds are shown in red in comparison with the full catalogue in the time–magnitude plots (e, k) and the maps (f, l) to highlight where these lie within the catalogue.

By examining the individual components of the smallest η_j in both of these catalogues, we find similarities in the distribution of parent magnitudes—the events with smallest η_j are most likely to be linked to small magnitude events, which explains our observation that the number of events at small nearest neighbour distances reduces at a higher magnitude cut-off. Though this observation holds for both catalogues it is also true that the smallest η_j events are not exclusively linked to small parents. A significant number of these events, especially in the Italian catalogue, are linked to larger magnitude parents which would result in smaller η_j due to the multiplicative nature of the rescaled distance (eq. 1).

The time to parent for small events in Italy (Fig. 6c) has a bimodality with a significant number of the events with small η_j having a small time to parent. The New Zealand small η_j time to parent distribution (Fig. 6i) does not demonstrate this bimodality with most of the small η_j events occurring at larger time to parent and only a few events at the shortest times to parent.

The time–magnitude graphs (Figs 6 e and k) show us that the events with small η_j occur throughout the catalogue, with more of these events around the larger events in the sequence. We also see from the maps (Figs 6 f and l) that such events are not limited by geographic area and that there are simply more of them within the large earthquake sequences. In particular, it appears from the geographical location of these events within the Italian catalogue that the smallest η_j event pairs may be related to large earthquake sequences or swarms.

Though both catalogues have some events at larger distance to parent, there are many small η_j event pairs that have a very small spatial distance to parent. In this case these events are detected as occurring at the same location, and a small fixed spatial separation is then set by the algorithm. These are shown in grey in Figs 6(b) and (h). Thus we see that many events in the Italian catalogue occur at the same spatial location as their parents, and these make up more than half of the events with very small η_j. However there are also events that occur at spatial distances of $10^{-2}-10^{2}$ km, which would not be considered extraordinary. These events have a small total rescaled distance η_j due to (a) their time to parent, which can be as small as 10^{-10} years or (b) to the magnitude rescaling resulting from a large magnitude parent or (c) some combination of these factors due to the multiplicative nature of the total η_j.

Similarly in the New Zealand catalogue we see that events with small η_j most likely arise as a consequence of colocated events (Fig. 6h). Some of these events can be linked with mining or quarrying operations and others are related again to large earthquake sequences, especially the 2016 Kaikoura earthquake. There is a marked increase in the number of events with small η_j from 2012 onwards, when GeoNet switched to calculating locations automatically. This highlights potential issues with artefacts of data collection propagating into possible contamination of inference in underlying physical processes.

The New Zealand and Italian catalogues allow us to explore the consequences of colocated events on the method (e.g from quarry blasts or repeating events identified and located using template matching). In such cases the rescaled interevent distance becomes 0 and therefore $\eta_j \equiv 0$. This introduces a problem for the (log)normal mixture, since we cannot take the log of 0, but is also problematic as events will be linked to this parent due to the very small space component regardless of the time component. The choices on how to proceed are (i) to remove all of the colocated events (possibly

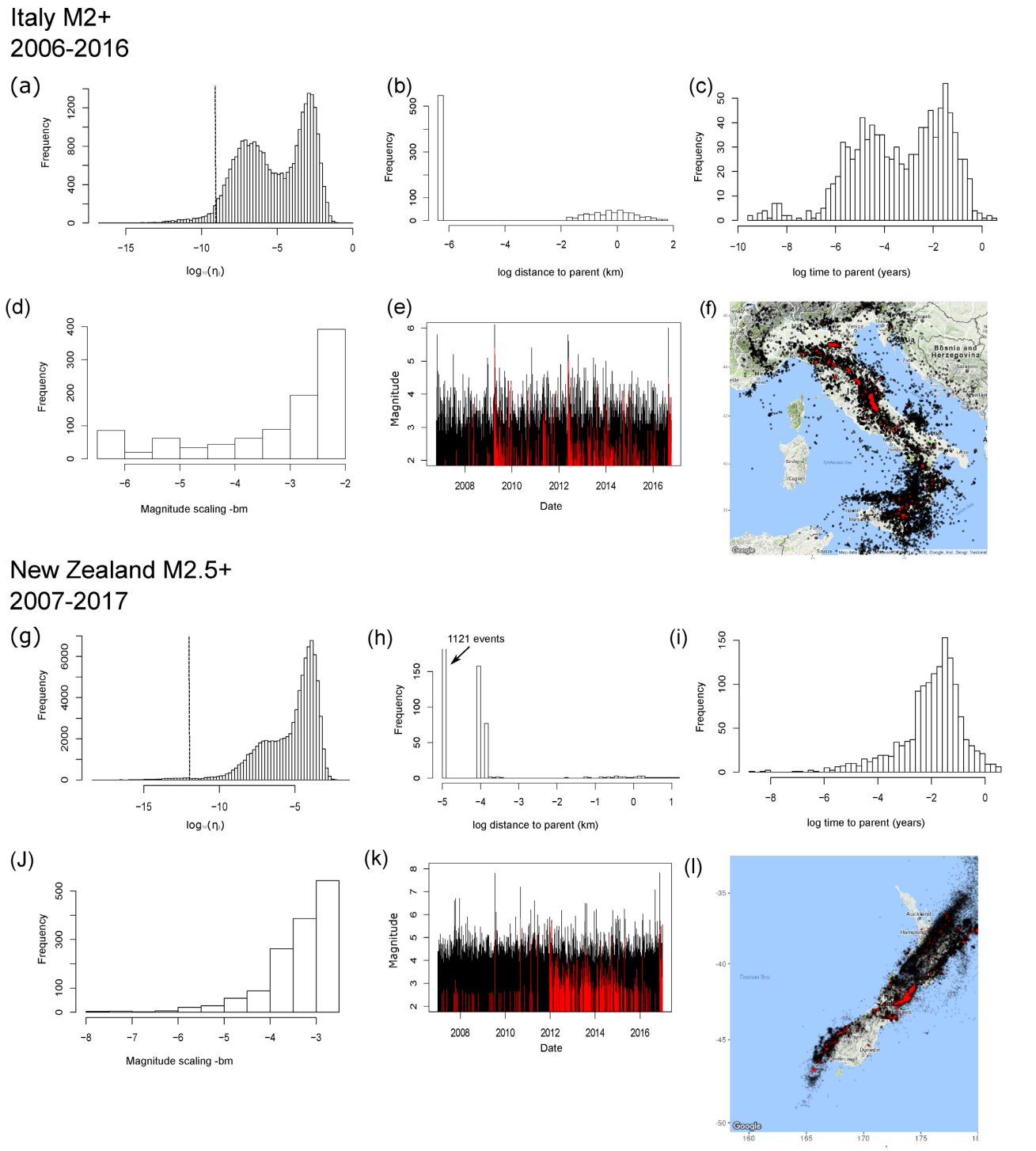

Figure 6. A comparison of the smallest η_j events in the Italy and New Zealand catalogues. The η_j distribution (a, g) for both catalogues includes a significant number events at very small η_j, those to the left of the dashed line. The smallest η_j can be deconstructed into the individual components—(nonrescaled, spatial) distance (b, h, with colocated events given a small fixed separation shown in grey), time (c, i) and b-value scaled parent magnitude (d, j) components. These small η_j events are shown in red in the time-magnitude (e, k) and location (f, l) plots of each catalogue.

retaining the first, but recognizing that there may well be a preceding event beyond the applied temporal window), (ii) not to allow connections between colocated events or (iii) to set the distance between colocated events to a small finite value such as the location uncertainty. All of these choices have problems from a theoretical perspective, particularly since true repeating events may genuinely be colocated. We take the pragmatic choice of setting the spatial distance to some minimum value. The consequence of this is that we could treat these colocated events as a third component; rather than doing so here, we leave this as a discussion point to address in future work.

3.3 Probabilistic cluster networks

Once the mixture model components have been modelled, these distributions can be used to identify individual earthquake clusters. In graph theory, a spanning tree is a graph which links all vertices of a graph using the smallest number of links between events. In this case, a spanning network can be generated by linking every event in the catalogue to its nearest neighbour parent only. In order to extract families of related events it is then necessary to crop the links between events believed to be independent, leaving the clustered events linked. Zaliapin & Ben-Zion (2013a,b) used a binary threshold to deterministically separate the links into two populations termed 'background' and 'clustered' events and trimmed the spanning network to retain the clustered linkages. They then went on to analyse the properties of these remaining clustered networks. This process does not appropriately handle uncertainty resulting from the overlapping distributions of background and clustered events illustrated in Sections 3.1 and 3.2.

We improve on this method through the construction of probabilistic cluster networks, which propagate uncertainty in the links through to the resulting cluster sequences. Once a mixture model has been fitted to the rescaled nearest neighbour distances, the probability that an event with a given η_j belongs to the clustered component can be calculated as shown by the coloured curves on the right hand column of Fig. 2. This allows us to assign a probability of linkage for each event to its nearest neighbour parent, based on the distribution the η_j most likely belongs to. The links in the network can then be 'thinned' in a similar manner to the stochastic declustering method of Zhuang *et al.* (2002, 2004), where the clustering probability can be compared with a random variable to decide if the link should be retained or not. The use of the MCMC approach for the mixture modelling allows us to repeatedly resample from the parameter posterior distribution when assessing the uncertainty in a given link. This stochastic thinning removes links from events which are identified as background and results in a number of cluster networks within the catalogue. Each link in the catalogue has some weight or certainty calculated by how often it is retained over many stochastic realizations, which is determined by the conditional probability of an event belonging to the clustered component, using different samples from the MCMC chain for each stochastic realization.

This method makes use of the full distributions identified in the MCMC mixture model to condition which links should remain and thus explicitly takes into account the overlap of the two distributions. The use of the MCMC method also allows us to quantify the uncertainty in the retained links and reduce the likelihood of miscategorizing events. Uncertainly linked events could then be studied further by incorporating physical information into the analysis or by making use of reported errors in location which are known to cause misclassification of events.

In Fig. 7, we show two examples of these probabilistic cluster networks constructed using our new method. These networks are included as a demonstration of how the probabilistic approach could be used to study cluster sequences, using two examples of structurally different sequences to highlight what can be gained from this approach. We identify which links remain stochastically and then colour the links based on their certainty, such that black links exist in 75 per cent or more of realizations and yellow links in less than 25 per cent, with green and blue representing probabilities in the range 25–50 per cent and 50–75 per cent, respectively. We have also assigned symbols to illustrate the magnitude of each event within the cluster and coloured the events by the time at which they occur relative to the mainshock, defined here as the largest event in the sequence. The earliest events, including any foreshocks, are coloured light pink, then red and grey with events at more than 1 yr after the mainshock in black. This colouring is consistent across all of the graphs in the figure. This allows us to examine the structure of individual clusters and identify where the most uncertain links exist. Fig. 7 includes examples from two different clusters in the Southern California catalogue for a minimum magnitude cut-off of 2.6. Zaliapin & Ben-Zion (2013b) suggest that the structure of the network can be used to classify a cluster as mainshock–aftershock or swarm type sequence, so we take an example of each as explored in their paper.

Fig. 7(a) is the probabilistic network for the 1981 Westmorland sequence in the Salton trough and 7(b) the Landers sequence. The Landers sequence is more burst-like in nature with one large event (M7.3) linked to many smaller events in a fairly traditional aftershock sequence where some of the larger aftershocks have child events of their own. The probability of an event being linked decreases with time from the mainshock, so while events are occasionally linked at more than a year after the mainshock, the probability that such an event would be linked is low. For the Westmorland sequence the events are linked more linearly forming a chain rather than a tree. Though there is one large event (M5.75) with several aftershocks, most of the events in the sequence are of magnitude $\leq$4.

Using these networks we plot the time variation of the cumulative number of events (c, d), the frequency–magnitude distribution (e, f) and spatial (g, h) distribution of individual clusters to identify features of interest. Figs 7(c) and (d) show the development of the cluster in time from 5 days before the mainshock to 10 days after. The burst-like structure of the Landers sequence and the large magnitude results in a smooth increase in the cumulative number of events with time, while the more 'swarmy' Westmorland sequence grows in a less constant manner with several events preceding the largest event. We call this a 'mainshock' in retrospect, though it may not have the same significance as it would in a 'mainshock–aftershock' sequence. The Westmorland sequence frequency–magnitude distribution (Fig. 7e) contains many small magnitude events and one large event with few intermediate magnitude aftershocks while the Landers sequence (Fig. 7f) has a frequency–magnitude distribution that is more consistent with a Gutenberg–Richter distribution. We also note that these sequences can be markedly different depending on the magnitude cut-off chosen. The mainshock magnitude of the Landers event is large enough that increasing the threshold magnitude does not have a large affect on the cluster structure; simply removing the smaller events keeps the overall structure intact as it is focused around the mainshock. For the Westmorland sequence, removing the smaller events significantly reduces the size of the cluster as small events are more fundamental to the generation of the more linear structure. This suggests that a careful choice of magnitude cut-off would be needed to study swarm-type structures in particular, where the events are often small in magnitude. This also suggests that missing small events from the catalogue could have a significant impact on observed clustering. The effect of missing events on resulting clusters has been studied using the ETAS model by several authors (Zhuang *et al.* 2013; Wang *et al.* 2010; Seif *et al.* 2017) and care should be taken with this in future seismicity analysis.

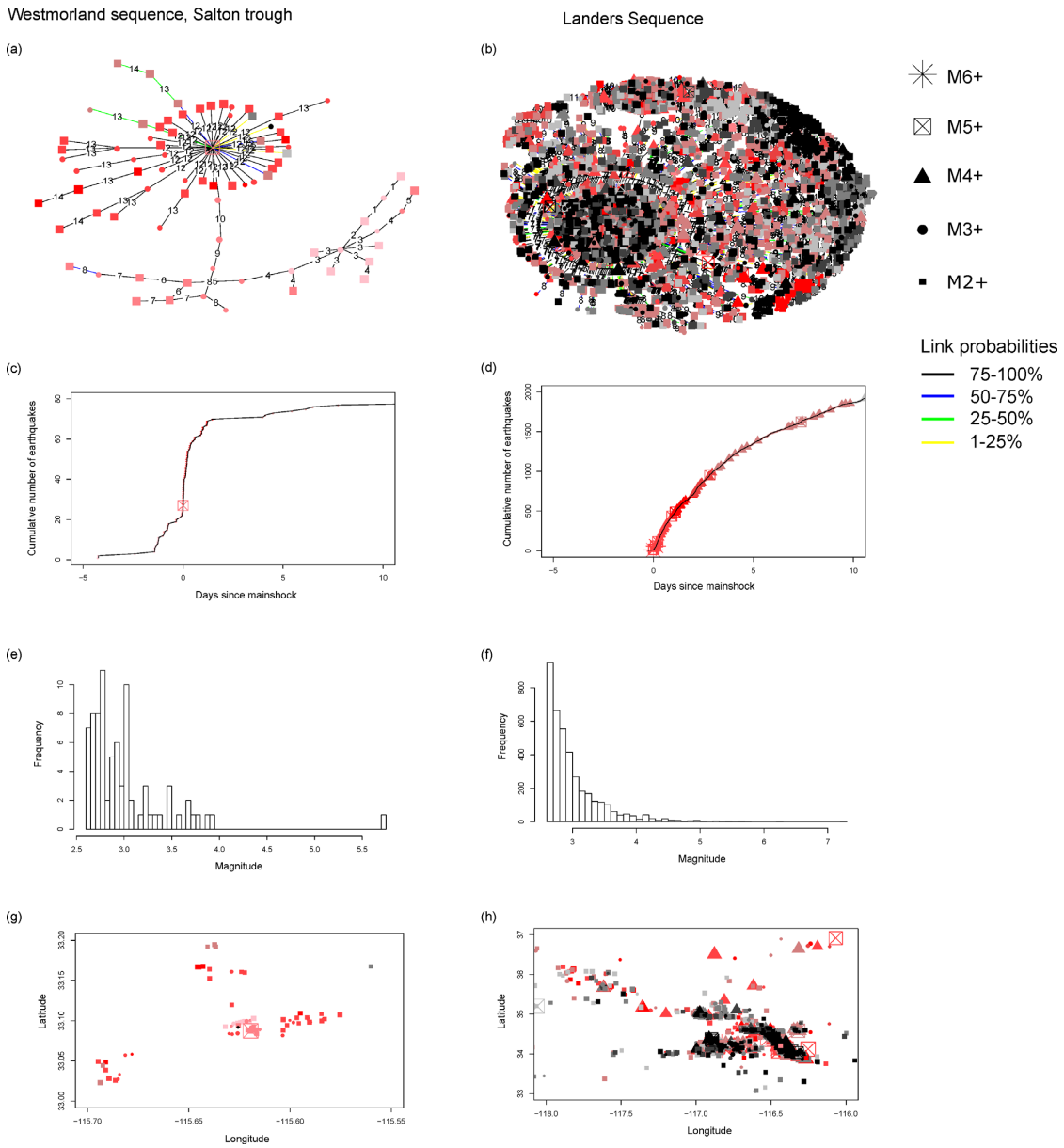

Figure 7. Clusters identified from the Southern California catalogue with minimum magnitude 2.6. Probabilistic network for (a) Westmorland sequence with mainshock magnitude 5.75 and (b) Landers earthquake sequence, mainshock magnitude 7.3. Cumulative events with time (c, d) for each sequence, magnitude distribution (e, f) of each cluster and spatial location (g, h) of events are also shown with the same colour/shape allocation used throughout. The shape and size of the marker is determined by the event magnitude (see key). The events are coloured by time relative to the mainshock, with any events before the mainshock coloured pink and events occurring more than 100 d after the mainshock in black with red events closest to mainshock and grey events between 10 and 100 d after the mainshock. The links between events are coloured by their certainty with black events $\geq$ 75 per cent likely to be linked, blue links between 50 and 75 per cent certain, green links 25–50 per cent certain and yellow links $\leq$ 25 per cent certain.

3.4 Leaf depth uncertainty

Zaliapin & Ben-Zion (2013b) propose a topological average leaf depth metric to discriminate swarm and mainshock–aftershock sequences. The average leaf depth is an average of the depth, or number of links, from the first event in a sequence to each leaf, where a leaf is an event with no offspring of its own. The idea is that swarm type sequences will then have a higher average leaf depth due to their more linear linkage while mainshock–aftershock type sequences will have a lower average leaf depth due to the greater number of first (or second) generation offspring events around a large mainshock. We investigate the robustness of such a metric by examining how it changes when considering the link uncertainties. We have chosen to focus on the average leaf depth as this metric is quite intuitive to understand and provides valuable information on the cluster structure which is not so easily achieved from other clustering metrics. To examine how the link uncertainties affect the average leaf depth, we take average leaf depths calculated from the results of 100 stochastic realizations. This involves resampling from the MCMC posteriors 100 times, calculating the cluster probabilities each time for a slightly different mixture fit and cutting links based on comparison with a random variable. For each realization we have automatically calculated the leaf depth, cluster size, the number of leaves—defined as events with no offspring—and other

such statistics for each cluster. Using the Southern California catalogue (Hauksson *et al.* 2012) for events with magnitude ≥ 2 (a magnitude cut-off of 2), the number of clusters identified in a realization varies from 7930 clusters to 8159 clusters, where a cluster is a sequence of at least two linked events. If there were no uncertainty in the method of cluster construction the number of identified clusters would not change.

To demonstrate how the link uncertainty is reflected in the clusters themselves, in Fig. 8(a) we plot average leaf depth against the number of leaves (events with no offspring) for five example sequences over 100 realizations—the Landers and Westmorland sequences discussed earlier, the 2005 Obsidian Buttes swarm (mainshock magnitude 5.1), the 1990 San Gabriel earthquake (mainshock magnitude 5.51) and a swarm in the Coso region in January 1982 (mainshock magnitude 5.12). Zaliapin & Ben-Zion (2013b) suggest that considering $M2+$ clusters with mainshock magnitude >4 and family size of ≥ 10, there is a bimodality in the average leaf depth variation with cluster size and number of leaves. In Fig. 8(a) the dashed line represents the boundary suggested by Zaliapin & Ben-Zion (2013b) of average leaf depth $=5$. We plot the boundary here to allow comparison between average leaf depths obtained by the two differing methods. Our chosen examples also fit the criteria of cluster size and mainshock magnitude.

Fig. 8(a) shows that the more 'swarmy' sequences (Coso, Westmorland and Obsidian Buttes) have significant variation in average leaf depth across this boundary, such that in some realizations they have leaf depths around 1 and in other realizations calculated leaf depths of more than 15. The Obsidian Buttes swarm sequence also demonstrates significant variation in the number of leaves, though the Westmorland and Coso sequences appear to have a more stable number of leaves and overlap significantly with each other. We might expect the swarm-type sequences to have a smaller number of leaves due to the more linear linkages in the cluster structure, but as demonstrated in Fig. 7(a), swarm-type sequences at a low-enough magnitude cut-off can display significant branching. The San Gabriel sequence is a typical mainshock–aftershock or burst-like sequence with one mainshock generating many offspring, and demonstrates less uncertainty in average leaf depth. The Landers sequence has a very high average leaf depth in many realizations, though this is not always the case as some realizations return an average leaf depth similar to that of the smaller San Gabriel sequence. Zaliapin & Ben-Zion (2013b) report the average leaf depth of the Landers sequence as 11.6, which is higher than the average leaf depth we calculate for this sequence in any of 100 realizations, and again we see variation across this boundary.

In mainshock–aftershock sequences, the average leaf depth can be increased significantly by events before the mainshock. Linking even one or two smaller foreshock events will increase the average leaf depth of the cluster. In Fig. 8(b) we plot the number of foreshocks against the average leaf depth, where we have defined foreshocks as any event within the cluster that occurs before the mainshock in time. At low number of foreshocks, we see that the leaf depth increases as the number of foreshocks increases. For the Landers and San Gabriel sequences we can see that the leaf depth increases almost linearly with the number of foreshocks, with the solid black line showing a 1:1 increase in leaf depth with increasing number of foreshocks. For large sequences, the leaf depth can be increased significantly with a few foreshocks. If there is an event directly before the mainshock, the leaf depth of all branches increases. This appears to flatten off at high number of foreshocks in the case of the Landers sequence. There are two realizations in which the San Gabriel sequence has 29 or 30 foreshocks, which is significantly higher than the <5 foreshocks identified in most realizations. Despite this significant number of foreshocks, the average leaf depth for these realizations is still <5. In these realizations, as in the realizations with a large number of foreshocks for the Landers sequence, the foreshocks occur on alternate branches of the cluster from the mainshock and are linked to the mainshock via a branching event. This increases the number of identified foreshocks and the number of leaves without increasing the average leaf depth significantly, as the leaf depth of the mainshock itself is not affected.

Though the variation of average leaf depth with foreshock activity can also explain some of the variation in the more 'swarmy' sequences, there is still significantly more variability than can be accounted for in an increased number of foreshocks, and the relationship is less linear as shown in Fig. 8(b). This is related to the more linear nature of linkages in swarm type cluster sequences, because cutting one link can drastically reduce the cluster size, where cutting a link in a mainshock–aftershock sequence will most likely only reduce the cluster size by a few events. Similarly, due to the linear nature of these sequences, cutting one link can also reduce the number of leaves if a cropped branch leads to an event with many offspring. This explains the variability of the number of leaves in the Obsidian Buttes sequence. This is less of an issue for the Westmorland sequence because almost all of the events before the mainshock are smaller than magnitude 4 and less likely to have identifiable offspring of their own above the magnitude cut-off. Both the Westmorland and Coso sequences also contain a more obvious mainshock event than in the Obsidian Buttes swarm, which may contribute to their more stable number of leaves.

The average leaf depth is therefore a measure of the linearity of the sequence before the mainshock. For sequences which are primarily mainshock–aftershock where there is less branching of the sequence, this results in an increased average leaf depth where there are a significant number of foreshocks. For more 'swarmy' sequences there is more variability because the number of events before the mainshock is often large and each of these events will have the possibility to generate offspring of its own. Swarm-type sequences demonstrate more variation in leaf depth due to their more linear nature—foreshocks are more likely to be cropped in a swarm sequence because there are more of them. This is true in the examples we have highlighted, but also highlights an issue with identifying swarm sequences which lack an obvious mainshock. The above analysis is based on identifying clusters by their mainshock, though we find this is not ideal for swarm type sequences where events are of similar magnitude and therefore the mainshock can change between realizations. The average leaf depth for our example swarms is high because the mainshock is located at the end of a linear chain of events and by choosing examples with this feature we are biasing our interpretation of swarms.

Furthermore, by defining swarms as events with high leaf depths and thus as long chains of events, any sequence with a significant number of events before the mainshock could be classed as a swarm. Instead of defining a bimodal split when considering all cluster sequences, we see a continuum, with many sequences having a range of average leaf depths dictated by the fraction of their population directly linked to the mainshock and the number of events that occur before the mainshock. We can also see significant variability in the average leaf depth as a reflection of link uncertainty. In some realizations multiple smaller clusters will be linked together, such as the small burst sequence that is included in two realizations of the San Gabriel sequence. The average leaf depth is not sensitive to this, but this highlights the new information that can be identified by including uncertainty in our analysis. The propagation of

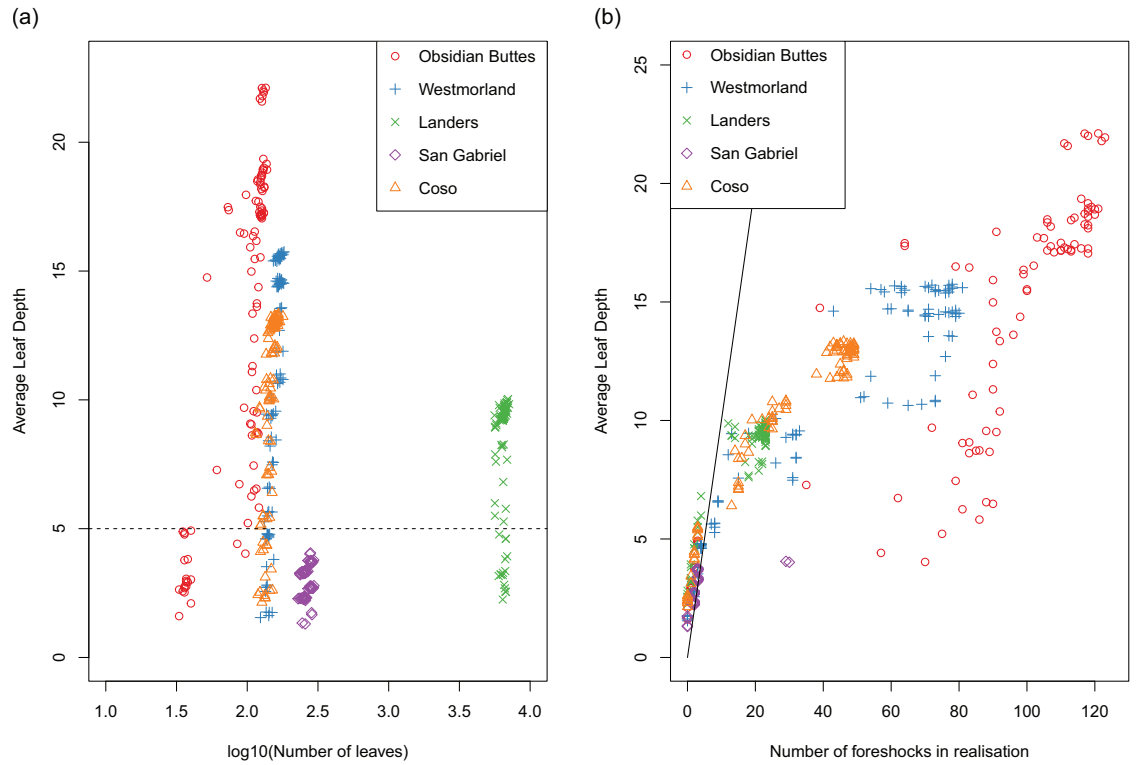

Figure 8. Variation in average leaf depth over 100 realizations of (a) the Obsidian Buttes, Westmorland, Landers, San Gabriel and Coso sequences against number of leaves, defined as events with no offspring. The dashed line marks the leaf depth discussed by Zaliapin & Ben-Zion (2013b) at which sequences can be classed as either mainshock–aftershock or swarm type clusters. Average leaf depth variation with number of foreshocks for each cluster is shown in (b). The solid line represents how average leaf depth would change with number of foreshocks if the relationship were 1:1.

uncertainty into cluster metrics may provide greater insight into the behaviour of different types of earthquake sequences and prove helpful for establishing robust methods of discriminating between different cluster styles.

From this we conclude that while the average leaf depth metric is indeed helpful for characterizing the structure of a sequence, the uncertainty in the average leaf depth and the mechanisms by which the average leaf depth can be increased make it less useful as a solitary metric of cluster type. The uncertainties in average leaf depth over many realizations are not unique to this metric, but simply a consequence of the underlying uncertainty in classifying events as background or clustered, and we argue that these uncertainties can provide insight into the clusters themselves, as we have discussed here for the average leaf depth. The metric is also particularly unhelpful for identifying swarm sequences, which will often lack the prominent mainshock that is such a vital component of the average leaf depth metric.

4 DISCUSSION

4.1 The problems with a discrete partitioning of the clustered and background components

As described in the introduction, it is common practice to select a binary threshold rescaled distance η_j below which event pairs are categorized as clustered and above which events are assumed to be background. The methods used to choose the threshold have been either a subjective identification of the minimum between the two peaks (e.g. Zaliapin *et al.* 2008) or using normal mixture modelling to identify the intersection point of the two distributions (e.g. Zaliapin & Ben-Zion 2013a, 2016a). It is important to understand the consequences of this binary thresholding.

Although this approach is convenient, the overlap between the clustered and background distributions means that a choice of a single hard threshold will inevitably mean that some clustered events pairs will reside in the region classified as background and some background event pairs will reside in the clustered part. This is true unless the background and clustered components are entirely separate, which is not commonly a case of interest.

The impact of an assumption of a binary threshold is illustrated in Figs 2(e) and (f). Since the Weibull distributions are asymmetric, there is asymmetry in the misclassification—specifically, the tail of the background component stretches significantly towards short distances. Using the mean parameter values the estimated number of misclassified events resulting from the use of a binary threshold can be calculated for a synthetic data set derived from the Weibull fit parameters. Using the threshold proposed by Zaliapin & Ben-Zion (2013a) of $\eta \approx 10^{-5}$ for the same Southern California data set, almost 24 per cent of background events are included in the clustered distribution, while as few as 0.5 per cent of clustered events are wrongly identified as background. This results in almost 12 per cent of total events being wrongly classified in the binary threshold system. This would lead to an underestimation of the background rate.

We conclude that the use of a binary threshold is a suboptimal approach for determining background rates if the clustered and background event distances overlap significantly. In this case a significant number of background events would be wrongly removed from the catalogue of independent events. Here we advocate

a different approach—acknowledging the overlap and assigning a probability that nearest-neighbour pairs are causally linked. Even so, some misclassification is possible. For example, in some cases background rates may be overestimated by a probabilistic approach due to spatially and temporally overlapping aftershock sequences masking the true background rate (Touati *et al.* 2011, 2014).

4.2 Probabilistic approach

We acknowledge that the probabilistic approach outlined here also has its drawbacks. The MCMC algorithm is more time consuming than the simpler maximum likelihood approach, but the MCMC approach allows a quantification of the uncertainty in the fits and the stochastic step allows some quantification of uncertainty in event allocation to either the clustered or background component. This allocation uncertainty will always exist in a mixture model framework due to the overlap of the two distributions, unless the two components are sufficiently separated. This uncertainty in allocation is not considered in a binary framework. In a comparison of the abilities of different algorithms to identify clustering, Sornette & Utkin (2009) compared the stochastic algorithms of Zhuang *et al.* (2002) and Marsan & Lengliné (2008) and suggested that a fundamental problem with the use of stochastic methods was that they resulted in many different possible models with no indication of which was the 'real' one. In this work, we demonstrate how the 'inherent uncertainty' in the clustering process described by Sornette & Utkin (2009) can be used to explore the relationship between events by constructing probabilistic cluster networks where each link reflects the propagated uncertainty, considering the uncertainty in the fits by sampling different values from the MCMC chain and the uncertainty resulting from the overlap of the distributions with the inclusion of the probabilistic thinning. This uncertainty clearly affects any cluster metrics used to discriminate between mainshock–aftershock and swarm type clustering (Section 3.4) and should therefore be considered in any methods for the discrimination of these different types of sequences.

A full catalogue analysis could make use of these uncertainties to investigate the underlying physical processes. In this case, clusters of events will consist of some strongly and some weakly linked events, and there may be other physical data that could relate to the link strengths. In examining cluster properties, the distribution of cluster metrics over many realizations may be helpful for constructing a robust method of discriminating between cluster styles. Current cluster metrics are essentially based on one realization of a cluster, where a probabilistic result will result in a distribution of cluster metrics that reflects uncertainty in the links between events.

Given a data set of around 10 000 events, the whole process from identifying nearest neighbours to sampling many realizations of the resulting clusters can be run on a standard desktop computer in under a day. For larger data sets, such as that of the 111 196 $M2+$ events in the Southern California catalogue, this would currently take a few days, but there is room for optimization which could significantly reduce this time.

4.3 Issues with colocated events and location uncertainty

As discussed in Section 3.2, colocated events occur in several catalogues and pose a problem for fitting mixture models. This may be particularly problematic when using catalogues supplemented with template-matching (e.g. Shelly *et al.* 2016) and therefore a serious consideration for future work as computationally generated catalogues become more common. Events at the same location can be set artificially to a very small spatial separation as in our algorithm, but this will cause a significant stretching of the clustered η_j mode, which will then have an affect on the mixture model fit by increasing the length of the tails of the distribution and/or adding a third mode. Though the locations are the most uncertain component of the η_j, Vasylkivska & Huerta (2017) suggest that the rescaled spatial distance is critically important in some areas, particularly those which experience induced seismicity. Care should be taken when dealing with small η_j events in the catalogue by identifying if the events can reasonably be removed, or if a small perturbation in location should be applied to reposition all events to account for location uncertainty.

Zaliapin & Ben-Zion (2015) also investigated the effects of location uncertainty in three Southern California catalogues. Though they did not report any issues with colocated events in these catalogues, they found that uncertainty in event location could have a significant effect on estimates of background rate, offspring productivity and fluctuations in *b*-value. They also highlighted how uncertainty in event locations may contribute to increased distance to parent, as mislocated events may find an alternate parent.

In cases where we are reasonably confident that the events do exist but are not necessarily actually at the same location, we suggest a stochastic displacement of such events so that their location is perturbed by some small amount within the location uncertainty, as opposed to the fixed approach used Section 3.2. If we then run several nearest neighbour fits, we can determine if it is likely that the parent is correct by how often the parent event changes. This approach can be applied in cases where there is a significant degree of location uncertainty and may go some way to addressing the issues discussed in Zaliapin & Ben-Zion (2015).

5 CONCLUSIONS

The bimodal distribution of rescaled nearest neighbour distances η_j in earthquake catalogues can be modelled with a two component mixture model that fits the data and identifies the underlying background and clustered components. The overlapping mixture model is preferable to a binary threshold which neglects this overlap. The use of normal mixtures overemphasizes the clustered component at large distances. The application of a gamma mixture overemphasizes the background component at short distances. The Weibull mixture provides a better fit to both the background and clustered components, with well-behaved tails. The shape of the distribution is important because it could significantly impact the assignment of events to clusters, especially when used with the binary threshold. The appropriateness of the Weibull mixture has been demonstrated on several catalogues. A small number of events occur in a possible third mode at very small rescaled distances, some of which are colocated and assigned a small separation by the algorithm. These often occur in extended earthquake sequences associated with large events.

An MCMC approach to mixture model fitting allows the full range of possible model fits to be identified and utilized to quantify uncertainty. A probabilistic approach to cluster identification based on the resulting distributions allows the assignment of uncertainties in the linkage between events in a candidate cluster. Our method can be used to construct probabilistic networks of earthquake sequences, where each link can be assigned some probability based on repeated stochastic thinning of the links in the cluster network.

These networks allow us to investigate the spatio-temporal evolution of earthquake clusters while considering our certainty in the linked events. The cluster metrics described by Zaliapin & Ben-Zion (2013b) exhibit significant uncertainty in the average leaf depth over many stochastic realizations, especially for the swarm-type clustering. This may be problematic when attempting to discriminate between mainshock–aftershock and swarm-type clustering. These findings may be useful for future study of individual earthquake clusters, in identifying large-scale patterns of clustering, and for accounting for clusters in future hazard zoning, hence improving the assessment of stationary and time-dependent seismic hazard.

ACKNOWLEDGEMENTS

We would like to acknowledge the UK Engineering and Physical Sciences Research Council for funding KB on agrant number 1519006. We acknowledge the New Zealand GeoNet project and its sponsors EQC, GNS Science and LINZ, for providing data. We thank Ilya Zaliapin and an anonymous reviewer for their thoughtful and constructive comments.

REFERENCES

Alvarado, G.E. *et al.*, 2017. The new Central American seismic hazard zonation: Mutual consensus based on up to day seismotectonic framework, *Tectonophysics,* **721,** 462–476.

Baiesi, M. & Paczuski, M., 2004. Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E,* **69**(6), 8.

Baiesi, M. & Paczuski, M., 2005. Complex networks of earthquakes and aftershocks, *Nonlin. Process. Geophys.,* **12**(1), 1–11.

Bourne, S.J., Oates, S.J., Van Elk, J. & Doornhof, D., 2014. A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir, *J. geophys. Res.: Solid Earth,* **119**(12), 8991–9015.

Davidsen, J., Gu, C. & Baiesi, M., 2015. Generalized Omori-Utsu law for aftershock sequences in southern California, *Geophys. J. Int.,* **201**(2), 965–978.

Diebolt, J. & Robert, C.P., 1994. Estimation of finite mixture distributions through Bayesian sampling source, *J. R. Stat. Soc., B (Methodological),* **56**(2), 363–375.

Enescu, B., Hainzl, S. & Ben-Zion, Y., 2009. Correlations of seismicity patterns in Southern California with surface heat flow data, *Bull. seism. Soc. Am.,* **99**(6), 3114–3123.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Petrol. Geol.,* **26**(4), 525–535.

Gentili, S., Di Giovambattista, R. & Peresan, A., 2017. Seismic quiescence preceding the 2016 central Italy earthquakes, *Phys. Earth planet. Inter.,* **272,** 27–33.

Hauksson, E., Yang, W. & Shearer, P.M., 2012. Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. seism. Soc. Am.,* **102**(5), 2239–2244.

Hicks, A., 2011. *Clustering in Multidimensional Spaces with Applications to Statistical Analysis of Earthquake Clustering,* Proquest, Umi Dissertation Publishing.

Jordan, T.H. *et al.*, 2011. Operational earthquake forecasting: state of knowledge and guidelines for utilization, *Ann. Geophys.,* **54**(4), 319–391.

Kagan, Y.Y., 2007. Earthquake spatial distribution: the correlation dimension, *Geophys. J. Int.,* **168**(3), 1175–1194.

Kamer, Y. & Hiemer, S., 2015. Data-driven spatial b value estimation with applications to California seismicity: To b or not to b, *J. geophys. Res.: Solid Earth,* **120**(7), 5191–5214.

Lunn, D., Spiegelhalter, D., Thomas, A. & Best, N., 2009. The BUGS project: evolution, critique and future directions, *Stat. Med.,* **28**(25), 3049–3067.

Maghsoudi, S., Baró, J., Kent, A., Eaton, D. & Davidsen, J., 2018. Interevent triggering in microseismicity induced by hydraulic fracturing, *Bull. seism. Soc. Am.,* **108**(3A), 1133–1146.

Maghsoudi, S., Eaton, D.W. & Davidsen, J., 2016. Nontrivial clustering of microseismicity induced by hydraulic fracturing, *Geophys. Res. Lett.,* **43**(20), 672–10.

Marsan, D. & Lengliné, O., 2008. Extending earthquakes' reach through cascading, *Science (New York, N.Y.),* **319**(5866), 1076–1079.

Marín, J.M., Rodríguez-Bernal, M.T. & Wiper, M.P., 2005. Using Weibull mixture distributions to model heterogeneous survival data, *Commun. Stat. - Simulat. Comput.,* **34**(3), 673–684.

Meletti, C., Galadini, F., Valensise, G., Stucchi, M., Basili, R., Barba, S., Vannucci, G. & Boschi, E., 2008. A seismic source zone model for the seismic hazard assessment of the Italian territory, *Tectonophysics,* **450**(1-4), 85–108.

Mohammadi, A., Salehi-Rad, M.R. & Wit, E.C., 2013. Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service, *Comput. Stat.,* **28**(2), 683–700.

Moradpour, J., Hainzl, S. & Davidsen, J., 2014. Nontrivial decay of aftershock density with distance in Southern California, *J. geophys. Res.: Solid Earth,* **119**(7), 5518–5535.

Peresan, A. & Gentili, S., 2018. Seismic clusters analysis in Northeastern Italy by the nearest-neighbor approach, *Phys. Earth planet. Inter.,* **274,** 87–104.

R Core Team, 2018. *R: A Language and Environment for Statistical Computing,* R Foundation for Statistical Computing, Available at: https://www.R-project.org .

Reverso, T., Marsan, D. & Helmstetter, A., 2015. Detection and characterization of transient forcing episodes affecting earthquake activity in the Aleutian Arc system, *Earth planet. Sci. Lett.,* **412,** 25–34.

Ruhl, C.J., Abercrombie, R.E., Smith, K.D. & Zaliapin, I., 2016. Complex spatiotemporal evolution of the 2008 Mw4.9 Mogul earthquake swarm (Reno, Nevada): interplay of fluid and faulting, *J. geophys. Res.: Solid Earth,* **121**(11), 8196–8216.

Schoenball, M., Davatzes, N.C. & Glen, J.M.G., 2015. Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field, *Geophys. Res. Lett.,* **42**(15), 6221–6228.

Schoenball, M. & Ellsworth, W.L., 2017. A systematic assessment of the spatiotemporal evolution of fault activation through induced seismicity in Oklahoma and Southern Kansas, *J. geophys. Res.: Solid Earth,* **122**(12), 189–10.

Seif, S., Mignan, A., Zechar, J.D., Werner, M.J. & Wiemer, S., 2017. Estimating ETAS: the effects of truncation, missing data, and model assumptions, *J. geophys. Res.: Solid Earth,* **122**(1), 449–469.

Shelly, D.R., Ellsworth, W.L. & Hill, D.P., 2016. Fluid-faulting evolution in high definition: connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm, J. geophys. Res.: Solid Earth, **121**(3), 1776–1795.

Sornette, D. & Utkin, S., 2009. Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks, *Phys. Rev. E,* **79**(6).

Sturtz, S., Ligges, U. & Gelman, A., 2005. R2WinBUGS: a package for running WinBUGS from R, *J. Stat. Software,* **12**(3), 1–16.

Touati, S., Naylor, M. & Main, I.G., 2009. Origin and nonuniversality of the earthquake interevent time distribution, *Phys. Rev. Lett.,* **102**(16), 168501.

Touati, S., Naylor, M. & Main, I.G., 2014. Statistical modeling of the 1997-1998 Colfiorito earthquake sequence: locating a stationary solution within parameter uncertainty, *Bull. seism. Soc. Am.,* **104**(2), 885–897.

Touati, S., Naylor, M., Main, I.G. & Christie, M., 2011. Masking of earthquake triggering behavior by a high background rate and implications for epidemic-type aftershock sequence inversions, *J. geophys. Res.: Solid Earth,* **116**(3), B03304.

Utsu, T., Ogata, Y. & Matsu'ura, R.S., 1995. The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth,* **43,** 1–33.

Vasylkivska, V.S. & Huerta, N.J., 2017. Spatiotemporal distribution of Oklahoma earthquakes: exploring relationships using a nearest-neighbor approach, *J. geophys. Res.: Solid Earth,* **122**(7), 5395–5416.

Vidale, J.E. & Shearer, P.M., 2006. A survey of 71 earthquake bursts across Southern California: exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *J. geophys. Res.: Solid Earth,* **111**(B5).

Wang, Q., Jackson, D.D. & Zhuang, J., 2010. Missing links in earthquake clustering models, *Geophys. Res. Lett.,* **37**(21).

Warnes, G.R. *et al.*, 2015. gtools: various R programming tools, *J. Stat. Software,* **12**(3), 1695.

Zaliapin, I. & Ben-Zion, Y., 2013a. Earthquake clusters in southern California I: identification and stability, *J. geophys. Res.: Solid Earth,* **118**(6), 2847–2864.

Zaliapin, I. & Ben-Zion, Y., 2013b. Earthquake clusters in southern California II: classification and relation to physical properties of the crust, *J. geophys. Res.: Solid Earth,* **118**(6), 2865–2877.

Zaliapin, I. & Ben-Zion, Y., 2015. Artefacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California, *Geophys. J. Int.,* **202**(3), 1949–1968.

Zaliapin, I. & Ben-Zion, Y., 2016a. A global classification and characterization of earthquake clusters, *Geophys. J. Int.,* **207,** 608–634.

Zaliapin, I. & Ben-Zion, Y., 2016b. Discriminating characteristics of tectonic and human-induced seismicity, *Bull. seism. Soc. Am.,* **106**(3), 846–859.

Zaliapin, I., Gabrielov, A., Keilis-Borok, V. & Wong, H., 2008. Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.,* **101**(1).

Zhang, Q. & Shearer, P.M., 2016. A new method to identify earthquake swarms applied to seismicity near the San Jacinto Fault, California, *Geophys. J. Int.,* **205,** 995–1005.

Zhuang, J., Ogata, Y. & Vere-Jones, D., 2002. Stochastic declustering of space-time earthquake occurrences, *J. Am. Stat. Assoc.,* **97**(458), 369–380.

Zhuang, J., Ogata, Y. & Vere-Jones, D., 2004. Analyzing earthquake clustering features by using stochastic reconstruction, *J. geophys. Res.: Solid Earth,* **109**(5).

Zhuang, J., Werner, M.J. & Harte, D.S., 2013. Stability of earthquake clustering models: criticality and branching ratios, *Phys. Rev. E,* **88**(6).